Sensitivity to grammatical and sociophonetic variability in perception

KATIE K. DRAGER

University of Hawai'i at Mānoa

Abstract

Phonetic realizations vary depending on social characteristics of the speaker, and recent research provides evidence that individuals are sensitive to at least some of these sociophonetic relationships during perception (Strand and Johnson 1996; Hay et al. 2006). In addition to socially-conditioned variation, there is evidence that phonetic realizations in production vary depending on the grammatical function of a word (Plug 2005; Hay and Bresnan 2006), yet it is not known whether listeners can actively exploit this phonetic variation in speech perception. This paper reports on three perception experiments conducted to determine whether perceivers' sensitivity to fine phonetic detail can assist in extraction of both grammatical and social meaning from the signal.

1. Introduction

Sociolinguistic ethnographies provide insights into how individuals manipulate both linguistic and nonlinguistic variants to construct their identities and they provide a way to examine this socially-conditioned linguistic variation without relying on the predetermined categories (e.g. age, sex, and socio-economic status) that are traditionally employed in sociolinguistic research (Eckert 1989). Recent research using experimental methodologies has shown that social and phonetic information are linked in the mind; the social information that gets attributed to a speaker can be influenced by altering which phonetic variants perceivers are exposed to (Campbell-Kibler 2007). Despite the advantages of both, there is little work which combines experimental methods with an ethnographic approach (Hay and Drager 2007). Combining these methods provides the opportunity to have an ethnographically informed interpretation of experimental results. Additionally, experimental work can illuminate the degree to which the social meanings of certain variables are either locally or globally-constructed (Maegaard to appear) or could potentially supply information about the degree to which subjects are sensitive to social and phonetic trends observed during an ethnography.

JLP 1 (2010), 93–120
DOI 10.1515/LABPHON.2010.001
1836-6346/10/01–93

Phonetic variation can also be lexically-conditioned and appears to be dependent on how frequently the word is encountered (Zipf 1929; Bybee 2002; Munson and Solomon 2004); the more frequent the lexical item, the more likely it is to have a realization that is phonetically reduced. Recent work provides evidence that such effects are also observed among realisations of different lemmas that share a wordform (e.g. *time* and *thyme*) (Gahl 2008; Drager 2009b). It is possible that in some cases word-based frequency effects are related to contextual factors in the signal rather than the frequency of the lemma itself (Jurafsky et al. 2002); words that are more predictable given their context are more likely to be phonetically reduced. However, if listeners are able to identify a lemma based solely on phonetic information in the token, it would provide evidence that (a) the fine phonetic detail is stored in the mind, (b) it is indexed directly to information about the specific lemma, and (c) the phonetic variation that is attributed to lexical effects in production is not simply an artifact of information elsewhere in the signal (such as in the environmental context).

This paper reports on results from three speech perception experiments conducted as part of a yearlong ethnographic study of at all girls' high school in New Zealand. Results based on the girls' production indicate that realizations of the word *like* depend on a combination of the pragmatic function of the word and the social grouping of the speaker. Informed by these results from production, the perception experiments presented here examine the link between social, phonetic, and grammatical (lemma-based) information in the mind.

2. Previous research

Recent research has begun to shed light on the degree to which phonetic and social information are linked in the mind. Campbell-Kibler (2007) displays how listeners make consistent social judgments about a speaker based on that speaker's use of known sociolinguistic variables, as in *fishin'* versus *fishing*. Maegaard (to appear) provides evidence that these judgments by listeners vary depending on a combination of locally-constructed meanings and more widespread meanings in a society.

There is also evidence that social information attributed to a speaker influences perception of speech sounds produced by that speaker (Strand and Johnson 1996; Strand 1999; Hay et al. 2006). In production, females are more likely to have the focus of aperiodic energy higher for both /s/ and /ʃ/ than males. To test whether individuals are sensitive to this relationship between gender and phonetic detail during speech perception, Strand and Johnson (1996) played gender-ambiguous tokens from an auditory /s/-/ʃ/ continuum. Each token was matched with a visual stimulus of either a male or a female face and participants were asked to identify the phone in a forced-choice task. Participants in their experiment were more likely to perceive /ʃ/ when shown a female face than when shown a male face. This provides evidence that listeners can use their (not necessarily conscious) knowledge about sociophonetic relationships to help determine what sound they heard. Similarly, participants in New Zealand, where there is an ongoing merger between /iə/ and /eə/, appear to be sensitive to a speaker's age during perception of these variants. When a voice was matched with a photograph of a younger face, Hay et al. (2006) found that participants were less accurate at identifying distinct tokens of /iə/ and /eə/. Younger New Zealanders are more likely to have merged tokens of /iə/ and /eə/, and perceivers seem to be sensitive to this relationship between a variant and the age of a speaker most likely to produce that variant.

There is strong evidence that phonetic variables are lexically-conditioned (Zipf 1929; Bybee 2001; Munson and Solomon 2004) and that such effects can even be observed across lemmas that share a wordform (Plug 2005; Gahl 2008). Research suggests that some of this variation is linked to how predictable a word is given its position in a sentence (Jurafsky et al. 2002). Work by Gahl (2008) provides evidence that words of homophone pairs, such as *time* and *thyme*, have consistently different durations; high frequency words are more likely to have shorter durations than low frequency words. Plug (2005) has found that the different functions of the Dutch discourse particle *eigenlijk*, glossed as 'actually', 'in fact', and 'now that I'm thinking about it', can have differing degrees of phonetic reduction. Additionally, Hay and Bresnan (2006) found that realizations of the vowel in the word *give*, which is involved in a sound change in New Zealand English, depend on whether the token of *give* is used as a verb of transfer or as part of an idiom as in *to give a hand*. Taken together, these results suggest that during production, speakers access a stored representation of a semantically and syntactically defined entry, or lemma, as opposed to solely a word-form entry, or lexeme. These semantically/syntactically defined representations must be stored complete with acoustic/phonetic detail or must be indexed independently to an additional representation where this phonetic information is available. This presents a challenge for models of speech production where lemma and phonetic-based information are separated by a phonological level (cf. Jescheniak and Levelt 1994; Levelt et al. 1999). A model that is consistent with these findings is an exemplar model of speech production and perception where utterances are stored as separate acoustically-rich exemplars and are indexed to social information of the speaker (Johnson 1997; Pierrehumbert 2001, 2006). This model will be discussed further in Sections 3.1 and 5.

3. Speech production at Selwyn Girls' High

The results presented in this paper are part of a larger study investigating identity construction and lemma-based phonetic variation within the context of an all girls' high school in New Zealand. I conducted a year-long ethnography at the school, which I refer to using the pseudonym Selwyn Girls' High (SGH) Drager (2009b). I spent four days a week at SGH for the entirety of the school day, spending my time among the students in their final year of high school. Of the 150 girls in their final year, I became familiar with roughly 70 of them though I also knew most of the other girls by sight. Through being present during their social periods (e.g. lunchtime), taking part in social activities (e.g. the formal), and making myself available as someone who would readily listen to their stories of struggle and achievement, the relationship that developed between myself and the girls who are the focus of this study does not resemble the anonymous researcher-participant relationship typical of experimental methodologies.

There are a number of friendship groups among the girls. Those which are the focus of this study are listed in Table 1. Some of these groups (e.g. The PCs, The BBs) conform to similar norms to one another, thereby setting and perpetuating the school's norms. For example, girls in groups listed on the left drink alcohol, wear clothes from chain stores found throughout New Zealand, and on rainy days eat lunch in the common room (CR), a room that is set aside for all Year 13 students though only some girls choose to use it. These shared characteristics contribute to the girls' styles and they reflect the girls' shared stance that they are "normal" and that everyone at the school gets along (Drager 2009a). Girls in these groups will be referred to as common room (CR) girls.

In contrast, some of these groups reject the norms of the CR girls and they reject the norms in different ways. For example, the Real Teenagers are rebellious; they party every day on the weekend and late into the night. They date guys who are older than those dated by the CR girls and they think the CR girls are shallow for only dating guys that society deems as attractive. Another NCR group, The Christians, are at the other extreme; they do not drink or date boys and they wear clothes that are more conservative than those worn by girls in any other group. One of the few characteristics that is shared by all of

Table 1. *Common Room (CR) and Non-Common Room (NCR) Groups, in no particular order.*

CR	NCR
The PCs	Pasifika Group
The Sporty Girls	The Goths
The Trendy Alternatives	The Geeks
Rochelle's Group	Real Teenagers
Relaxed Group	Sonia's Group
The BBs	The Christians
	Cecily's Group
	Loners

these norm-rejecting groups is that they do not eat lunch in the CR; they are non-common room (NCR) girls. The wide variety of styles among the NCR girls is consistent with their shared stance that they are “different”, that being “normal” is not necessarily a desirable characteristic.

3.1. *Phonetic variation of ‘like’*

This binary division between those who view themselves as normal and those who view themselves as different is reflected in where they eat lunch on rainy days: in the common room (CR) or not (NCR). To investigate whether this social categorization is also evident in phonetic variation and to examine lemma-based phonetic variation among words that share a wordform, acoustic phonetic analysis was conducted on tokens of the word *like* from the girls’ speech.

The word *like* has a number of different functions (Romaine and Lange 1991; Anderson 2001; D’Arcy 2007). D’Arcy (2007) distinguishes between those that are traditionally grammatical, such as the lexical verb (1a) and adverb (1b), and those that are discursive, such as the discourse particle (1c), the discourse marker (1d), and the quotative (1e). The examples presented here are taken from the interviews conducted with the girls.

(1a) I quite *like* you. (Rose, CR)
(1b) But I don’t think she’s *like* that as much anymore. (Meredith, NCR)
(1c) I was just *like* singing. (Rochelle, CR)
(1d) *Like* it real cracks me up. (Emma, CR)
(1e) She was *like* whoa. (Isabelle, NCR)

There has been a considerable amount of work investigating sociolinguistic variation in the distribution of the different functions of *like* (e.g. Tagliamonte and Hudson 1999; Dailey-O’Cain 2000; Tagliamonte and D’Arcy 2007), yet there is little work examining phonetic differences between the different grammatical categories. The author knows of only one study that mentions the potential of phonetic variation of *like*. Hazen (2006) found that, for one speaker of Appalachian English, discursive functions of *like* have different phonetic realizations of the vowel than is found in other words containing that vowel. He did not investigate whether this speaker’s phonetic realizations of the discursive functions differ from her realizations of the traditionally grammatical functions nor whether there was a difference in realizations between the different discursive functions.

The phonetic analysis, which is presented in Drager (2009b) is conducted on 720 tokens of *like* extracted from interviews with 28 of the girls. The tokens are made up of different functions of the word *like*: the discourse particle, the quotative, and the two most common traditionally grammatical functions (lexical verb *like* and adverbial *like*). These particular functions are analyzed because they are highly frequent in the speech of all of the girls and all four functions usually appear phrase-medially (as opposed to the discourse marker, which appears phrase-initially and phrase-finally).

As potential predictors of function type, the analysis tested the duration of the /l/ and the vowel, the targets of the nucleus and offglide, the mean pitch of the vocalic segment, the preceding and following environment, and whether any voiced portion of the token was produced with creaky voice (referred to here as glottalization). Also tested was whether the /k/ was realized; the /k/ was marked as present unless there was neither auditory nor acoustic evidence of a velar closure or there was no evidence of a fricated /k/. Glottalization was not treated as a realization of /k/ because periods of irregular durations between vocal pulses were sometimes present even when there was a clear closure and release of the /k/.

The F1 and F2 values extracted were converted to the Bark scale and were used to calculate the Euclidean distance between the F1 and F2 values of a target's nucleus and offglide. This provides a gradient measurement of degree of monophthongization; the smaller the Euclidean distance, the more monophthongal the token. Speech rate was calculated automatically using ONZE Miner (Fromont and Hay 2008) for the five seconds and twenty seconds surrounding each token, and it was tested as a possible predictor in the models. The duration of each segment (e.g. the /l/) was also tested. Consonant durations were normalized for speech rate by dividing the consonant duration by the duration of the vowel, resulting in, for example, the /l/ to vowel duration ratio of a token. A larger /l/ to vowel duration ratio indicates a longer /l/ relative to the duration of the vowel.

The relationship between these phonetic characteristics, a token's function, and whether the speaker was a CR girl or not was tested using mixed effects models. The phonetic factors that differed significantly across the different functions of *like* are summarized in Table 2.

The lexical verb and the adverb are similar phonetically and have been collapsed into a single category of traditionally grammatical functions as opposed to discourse pragmatic functions. The results indicate that quotative *like* is

Table 2. *Summary of production results.*

factor	quote-dp	gram-quote	gram-dp
nucleus F2	no diff.	gram smaller F2	gram smaller F2
diphthong	quote more monoph.	quote more monoph.	no diff.
high pitch	quote higher	quote higher	no diff.
/l/:V duration	quote shorter	quote shorter	no diff.
/k/ & CR	CR = quote less [k] NCR = quote more [k]	no diff.	no diff.
/k/ & freq of use	no diff.	use more = quote less [k] use less = quote more [k]	no diff.

more likely to be monophthongized than the non-quotative functions of *like*, even when compared to the other discourse pragmatic function analyzed: the discourse particle. Quotative *like* is less likely to be glottalized than discourse particle *like* and the difference in glottalization between the lexical verb and the discourse particle is not significant. Quotative *like* is less likely to have a large ratio of /l/ to vowel duration. In other words, when compared to non-quotative functions of *like*, quotative *like* is more likely to have a shorter /l/ relative to the length of the vowel. The results also reveal two different interactions involving /k/-realization: one where it interacts with whether the speaker is classified as a CR girl or a NCR girl and one where it interacts with how often the speaker uses quotative *like* when producing a quotative. CR girls are more likely to realize the /k/ in discourse particle *like* and drop the /k/ in quotative *like* whereas NCR girls are more likely to drop the /k/ in discourse particle *like* and realize the /k/ in quotative *like*. Above and beyond this effect is the effect of speaker-specific probability of quotative *like*. When producing a quotative, girls who are more likely to use quotative *like* rather than one of the alternatives are more likely to drop the /k/ in quotative *like*, regardless of whether they are CR or NCR girls.

These results lend support to findings from Plug (2005), Hay and Bresnan (2006), and Gahl (2008) demonstrating lemma-conditioned phonetic variation. Furthermore, because the trends for /k/ realization are in different directions for two different social groups, it is unlikely that the patterns are due to other aspects of the linguistic environment. This provides evidence that lemma-based phonetic variation is not limited to a frequency effect but can have social motivation.

Socially-conditioned lemma-based variation supports an experience based model of speech production, such as an exemplar model. Different individuals at the school systematically differ in their realizations of the different functions of *like* presumably due to a combination of differences in their experiences and the active expression of their identity through language use (Drager 2009a, 2009b). In an exemplar model of speech production and perception, utterances are stored as separate phonetically-rich exemplars (Pierrehumbert 2001). These exemplars are indexed with contextual information, such as social characteristics of the speaker (Johnson 1997). Both production and perception are biased toward recently and frequently encountered variants. In a hybrid model, exemplars are also indexed with an abstracted phoneme level representation (Pierrehumbert 2006).

Exemplar Theory predicts that, during perception, individuals will be sensitive to the grammatically and socially conditioned phonetic variation observed during production. As described earlier, previous research has shown that vowel perception is influenced both by trends in the speech of the perceiver and by trends associated with social characteristics attributed to the speaker. In the perception of *like*, individuals should differ depending on trends in their own production as well as trends in the production of who they believe the speaker to be. It also predicts that there will be differences in perception depending on the amount of exposure a perceiver has to the different functions of *like* produced by a particular speaker as well as other individuals who share social characteristics with that speaker.

4. Perception of *like*

To determine the extent to which the girls are sensitive to the relationship between social, grammatical and phonetic variation, three speech perception experiments were conducted. The experiments involve related, forced-choice tasks in which short clips from spontaneous speech produced by the participants and their classmates are used as stimuli. In the first two experiments, participants are asked to identify the function of an auditory token of *like*, and in the third experiment, they are asked whether an auditory token of *like* was produced by a girl who eats lunch in the CR or not.

All tokens are from spontaneous speech recorded during informal interviews conducted at the school. They are spliced from the original signal at a zero crossing in the waveform and at the same point of segmentation as used in the production analysis. All tokens with the /k/ present also have the /k/ released, and the token ends after the release has completed (if followed by a pause) or immediately before the following segment begins (e.g. before the closure period of a stop or a shift in the focus of aperiodic energy if followed by a fricative). Phonetic characteristics inherent in the tokens are not modified in any way.

A total of 42 girls took part in the perception experiments. One subject did not complete the last two blocks of Experiment 3 due to time restrictions. The experiments were conducted on the school grounds, and participants received a chocolate bar in exchange for their time.

4.1. *Experiment 1*

4.1.1. *Method.* In Experiment 1, the question of interest is whether participants can use phonetic detail to distinguish between different functions of *like*. Participants are played two tokens for each question. Each of the tokens in a question are different functions of *like* produced by the same speaker. The voices of 7 different girls (4 CR, 3 NCR) are used. There are 10 questions (5 produced by CR girls, 5 produced by NCR girls) that compare the discourse particle with one of the traditionally grammatical functions (either the lexical verb or the adverb) and 20 questions (10 produced by CR girls, 10 produced by NCR girls) that compare the discourse particle with the quotative.

The aim of the experiment is to determine whether participants can use phonetic cues to identify a function of *like*. In order to narrow down which phonetic cues participants use and reduce the risk that some other, unaccounted for signal is responsible for any observed trends, auditory tokens with a variety of phonetic characteristics were chosen as stimuli rather than solely representative tokens. For example, roughly equal numbers of quotative and discourse particle stimuli are monophthongal although in production quotative *like* is much more likely to be monophthongal. Therefore, if participants are highly accurate on the task, they would not appear to be using the phonetic characteristics identified in the production data but some other, unidentified cue.

The tokens of *like* were extracted from the original signal along with one or two words from the preceding context. Whenever possible, the complete tokens for a given question are matched at the lexical level (e.g. *he's like* and *he's like*). However, due to the difficulty of finding lexically matched tokens that are different functions and are produced by the same girl, tokens for some questions are mismatched at the lexical level (e.g. *I was like* and *he was like.*) Previous work on quotative *like* conducted in New Zealand indicates that Pākehā (New Zealand European) English speakers are more likely to produce quotative *like* with the historical present (i.e. present tense morphology with a past temporal reference), as in *he is like*, than with the past tense, as in *he was like* and that it is also most likely to occur with the first person singular (e.g. *I was like*) (Buchstaller and D'Arcy 2009).[1] None of the experimental stimuli for a given question differ in both of these respects, but some differ in either tense or person. Using the relative likelihoods observed by Buchstaller and D'Arcy (2007), questions for which quotative *like* is more likely to occur with the context of the first token than the context of the second token are labeled as "likely preceding" and questions where it is less likely to occur with the context of the first token are labeled as "unlikely preceding". Questions where the contexts are matched at the lexical level are labeled as "matched preceding". Questions comparing the discourse particle and the traditionally grammatical functions are not labeled according to their preceding contexts because relevant work has yet to be conducted with speakers of New Zealand English. The quotative could not be compared with the traditionally grammatical functions in Experiment 1 due to the low number of tokens identified that had matching preceding contexts.

For the task, participants are asked to indicate which context on the answersheet they believe each token was taken from. An example question is shown in Figure 1. The contexts provided are not actual utterances from the interviews, and the contexts for each question begin with the same sound in order to avoid a potential perception bias resulting from any coarticulation present at the end of the spliced auditory signal. The order of tokens within a question number is designed so that half of the time participants hear the discourse

a) I was like . . .	b) I was like . . .	
. . . gonna go til I heard that.	(a)	(b)
. . .“Go and grab it.”	(a)	(b)

Figure 1. *Example question from Experiment 1, where two auditory tokens, (a) and (b), were matched with contexts written on an answersheet.*

particle first. The contexts on the answersheet are also pseudo-randomized, such that half of the time that a token of discourse particle *like* is played first, the discourse particle *like* context is listed first on the page.

There is no training session, and participants sometimes fail to answer the first question. Therefore, question one, which compares a quotative and a discourse particle produced by a CR girl, is presented again later in the experiment in order to ensure that participants respond to it. After participants respond to all tokens produced by all 7 girls, the questions are repeated in the same order as in the first half of the experiment. Identical contexts are shown on the page in the same order as in the first half of the experiment, and tokens for each question are played in the opposite order in an effort to reduce the risk of a response bias.

4.1.2. *Results.* The results are summarized in their raw form in Tables 3 and 4. Because each combination of tokens was played twice during the experiment with the tokens in the opposite order, the effect of different factor groups can be determined by comparing the two opposite factors (e.g. first token mean EucD and second token mean EucD). If the values are the same for the two opposite factors, there is no co-variation between the factors and the tendency to indicate that a token was a particular function. A difference between the factors of a single factor group suggests that participants may have used those phonetic cues to help identify the function of *like*.

The results to questions that compare quotative and discourse particle tokens, shown in Table 3, indicate that 60% of the questions where the first token was identified as the quotative had the quotative context listed first on the answersheet. The results also indicate that the preceding contextual information may have played a role; when the preceding context of the first token is more likely to occur with the quotative than the context of the second token, the first token was identified as the quotative for 55% of the 335 questions where the preceding context was mismatched and this tendency is slightly stronger for the CR girls.

More monophthongal tokens and tokens with a lower F2 target in the nucleus tended to be identified as the quotative. Additionally, tokens with a shorter /l/ to vowel duration ratio were more often identified as the quotative. The results also indicate that when only one of the tokens in a question has the /k/

Table 3. *Characteristics of quote–dp questions in Experiment 1 where the first token is identified as the quotative, by whether the participant is in a CR or a NCR group.*

feature questions comparing	CR girl quote–dp	NCR girl quot–dp	CR and NCR quote–p
total number speakers	23	19	42
total questions answered	916	774	1690
total 1st token labeled as quote	465	383	848
quote first on sheet	278	231	509
1st token's context more likely	110	73	183
1st and 2nd tokens' context matched	271	242	513
1st token's context less likely	84	68	152
1st token mean EucD	1.5930	1.5400	1.5690
2nd token mean EucD	1.6180	1.6720	1.6430
mean EucD diff. (Bark)	−0.02538	−0.13280	−0.07388
1st token mean nuc F2 (Bark)	11.25	11.19	11.23
2nd token mean nuc F2 (Bark)	11.49	11.45	11.47
mean nuc F2 diff. (Bark)	−0.2379	−0.2554	−0.2458
1st token mean duration ratio	0.33900	0.32670	0.33350
2nd token mean duration ratio	0.35020	0.34520	0.34790
mean duration ratio diff.	−0.01120	−0.01844	−0.01447
1st token [k] present, 2nd token [k] absent	93	84	177
1st token [k] absent, 2nd token [k] present	74	65	139
[k] present for both tokens	118	83	201
[k] absent for both tokens	180	151	331

present, both CR and NCR girls identified the token with the /k/ present as the quotative token.

Responses to questions comparing the discourse particle with one of the grammatical functions (either the lexical verb or the adverb) are shown in Table 4, with the number of times the first token was identified as the grammatical function listed. Roughly 78% of questions where the first token was identified as the grammatical function had the context with the grammatical function shown first on the answersheet.

In order to examine the effects of each factor within the context of effects from the other factors, two mixed effects models are fit to the data from Experiment 1 using R (R Development Core Team 2007). The first examines responses to questions that compare the quotative with the discourse particle and the second examines responses to questions that compare the discourse particle with one of the traditionally grammatical functions. A mixed effects model allows the inclusion of random effects, such as those resulting from participant-specific trends, in addition to fixed effects. Fixed effects are those that are expected to be generalizable to other comparable populations of speakers and subjects. These can include linguistic factors, such as phonological environment, and social factors, such as the age of the participant. In this

Table 4. *Characteristics of gram–dp questions in Experiment 1 where the first token is identified a grammatical function, by whether the participant is in a CR or a NCR group.*

feature questions comparing	CR gram–dp	NCR gram–p	CR and NCR gram–dp
total number speakers	23	19	42
total questions answered	435	371	806
total 1st token labeled as gram	235	197	432
gram first on sheet	179	158	337
1st token mean EucD	1.1450	1.1750	1.1580
2nd token mean EucD	2.1720	2.3090	2.2350
mean EucD diff. (Bark)	−1.0270	−1.1350	−1.0760
1st token mean nuc F2 (Bark)	11.610	11.680	11.640
2nd token mean nuc F2 (Bark)	11.180	11.150	11.170
mean nuc F2 diff. (Bark)	0.4210	0.5324	0.4718
1st token mean duration ratio	0.6032	0.5812	0.5918
2nd token mean duration ratio	0.5878	0.6098	0.5992
mean duration ratio diff.	0.0017	−0.004056	−0.0009248
1st token [k] present, 2nd token [k] absent	35	35	70
1st token [k] absent, 2nd token [k] present	29	28	57
[k] present for both tokens	90	69	159
[k] absent for both tokens	81	65	146

type of model, the degree to which an effect is robust is determined within the context of the other effects included in the model. Random effects are those that are not expected to be generalizable. These can include the individual participant or the question number. If the participant is included in a mixed effects model, each participant is assigned their own coefficient. This allows participants to differ randomly from each other, reducing the risk that a single participant will bias results (Baayen 2008:263–309).

In the first model, only responses to questions comprised of quotative and discourse particle *like* are included. Whether the first token was identified as the quotative is treated as the dependent variable. Identification of the first token as the quotative is modeled rather than accuracy on the task in order to determine whether phonetic cues are used to identify a lemma in a way that is consistent with trends from production and in a way that is independent of the actual function of the token. Participant and question number are included as random effects in the model. Table 5 shows the coefficients of the fixed effects in the model. Effects tested but not included in the final model are whether the stimulus has the /k/ present, whether the participant is in a CR group, and whether the participant is more likely to produce the /k/ in quotative *like* than in the discourse particle. Also tested were different combinations of phonetic cues in the first and second token (e.g. whether the first token is monophthongal and the second is diphthongal). Only factors reaching significance are

Table 5. *Experiment 1 coefficients of fixed effects for model comparing responses to the quotative and the discourse particle.*

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.8036	0.1337	6.009	<0.0001
quote second	−1.0213	0.1056	−9.669	<0.0001
matched preceding	−0.2847	0.1387	−2.053	0.04007
unlikely preceding	−0.4539	0.1706	−2.660	0.00781
duration ratio difference	−0.7200	0.2519	−2.858	0.00426

included in the model shown in Table 5. The table does not include random effects because they are included as control factors, not as test factors. No interactions between factors reach significance in the model.

Fixed effects included in the model are whether the quotative context is listed on the answersheet after the discourse particle context (quote second) and whether a question is labeled as "matched preceding", "likely preceding", or "unlikely preceding". Also included in the model is the difference in ratio of /l/ and vowel durations between the first and second tokens for a question (duration ratio).

Estimates shown are in log odds. For the non-gradient factors, the model assumes one of the options as a default. The table does not display the default factors because the estimated coefficients for these factors are equal to zero. Here, the defaults are that the quotative context is listed first on the page and the more likely context for the quotative is heard first. For continuous factors, such as duration ratio, the model assumes a value of zero for the default. The intercept estimate is the log likelihood of identifying the first token as the quotative given the default factors. When determining the likelihood of a binomial factor, the estimated coefficient for that factor is added to the estimated coefficient for the intercept. For a continuous factor, the likelihood is determined by multiplying the factor's estimated coefficient with the value for the token. The p-values in the table indicate the significance level within each factor of the listed item from the default item. For example, "matched preceding" predicts identification of the first token as the quotative at the 0.05 level when compared with questions labeled as "likely preceding".

Participants are significantly less likely to identify the first auditory token as the quotative if the quotative context is listed second on the answersheet (quote second) ($p < 0.0001$). This trend reflects an overall bias for participants to identify the first token heard with the first context on the sheet. Including this factor in the model allows for examination of other potential factors that influence responses; the model holds this as constant when testing effects of the other factors. Additionally, the experiment design controlled for this through counterbalancing the auditory stimuli.

Participants are less likely to identify the first token as the quotative if the first auditory token of *like* has an "unlikely preceding" context ($p < 0.01$) than if it has a "likely preceding" context, and responses to tokens that are matched for preceding context fall between the two mismatched question types. This is in the expected direction given the trends observed by Buchstaller and D'Arcy (2007). Participants are more likely to identify a token as the quotative if its preceding context indicates a higher probability of quotative *like* occuring relative to the other token with which it is matched. This finding provides evidence that individuals are sensitive to lemma-specific contextual information during perception.[2]

The difference in the /l/ to vowel duration ratio between the first and second tokens (duration ratio difference) reflects how much longer the duration of the /l/ in the first token is when compared to the duration in the second token; a positive value for duration ratio difference indicates that the /l/ to vowel duration ratio in the first token is longer than that in the second token. As indicated by the negative coefficient for this factor in Table 5, participants are less likely to identify the first token of quotative *like* if it has a larger duration ratio than the second token ($p < 0.01$). In production, quotative *like* is more likely to have a smaller ratio of /l/ to vowel duration than discourse particle *like*; /l/ is shorter in quotative *like* than in discourse particle *like*, relative to the duration of the vowel. Listeners' responses are consistent with trends in their production.

A second mixed effects model was fit to responses from Experiment 1 for questions that compare the discourse particle with one of the traditionally grammatical functions. It models the likelihood of the first token being identified as the traditionally grammatical function. Participant and question number are included as random effects. All potential predicting factors tested for the first model were also tested for the second with the exception of the likelihood of the preceding context, which was not tested due to lack of previous work investigating contexts of the traditionally grammatical functions in New Zealand English. The final model is shown in Table 6.

The only fixed effect that reaches significance in the model is whether the traditionally grammatical token is listed on the answersheet after the discourse particle token (gram second); participants are less likely to identify the first token as the grammatical function if the second context on the answersheet is the traditionally grammatical context ($p < 0.0001$). As with results from the

Table 6. *Experiment 1 coefficients of fixed effects for model comparing responses to the discourse particle and grammatical functions of* like.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.3933	0.1003	3.923	<0.0001
gram second	−0.8205	0.1789	−4.586	<0.0001

first model, this trend reflects a bias of the participants to identify the first token heard with the first context on the answersheet, regardless of phonetic or contextual cues in the auditory token.

This tendency to match the first token heard with the first context listed is the forced-choice matching task equivalent of an acquiescence response set (the tendency for participants to answer "yes" for yes/no questions in experimental work), an effect which is commonly found in the psychology literature (cf. Bentler et al. 1971).

In order to test participants' ability to identify a function independent of cues inherent in the preceding context, a second experiment was run in which the stimuli consisted only of the word *like*.

4.2. *Experiment 2*

4.2.1. *Method.* The design for Experiment 2 is similar to that in Experiment 1: participants are asked to match an auditory token containing *like* to the context they feel it is most likely to have come from. In contrast to the first experiment, the auditory tokens consist only of the word *like*. The voices of four girls from different groups at SGH are used and the experiment is blocked by voice.

In contrast to the longer clips in Experiment 1, the shorter clips in Experiment 2 allow for a three-way comparison between the different functions of *like*. Five of the tokens for each voice are grammatical functions of *like* (either a lexical verb or an adverb), five are quotative *like*, and five are discourse particle *like*. Participants are asked to distinguish between grammatical and quotative *like*, grammatical and discourse particle *like*, and quotative and discourse particle *like*. The two auditory tokens for each question number are produced by the same speaker, and stimuli are blocked for each voice. After responding to 15 questions for each voice, participants are asked whether they recognize the voice and, if so, who they think the speaker might be. If they think a voice sounds familiar but they are unsure of who the speaker is, they are told to indicate that they recognize the voice and they can either not name the speaker or else name multiple people who they think the speaker might be.

The contexts provided on the response sheet differ for each question within a single block. The same contexts are used across the different blocks, but they differ in the order they appear within a particular question and the order in which the context pairs are listed. For example, the contexts for speaker 1, question 3 are in the following order: *I was like "Only if he asks me himself"* and *I was like only two seconds behind*, whereas they are in the opposite order for speaker 2, question 21. As in Experiment 1, participants are told that the contexts are not the actual contexts from the interview but that they are similar. The manner in which they are similar is not made explicit.

After playing stimuli for all four voices, the first half of the experiment is repeated. The questions are presented in the same order as during the first half, but the order in which the auditory tokens are played within each question is reversed in order to counterbalance potential effects from a response bias based on the tokens' order. The contexts are presented in the same order as found in the first half of the experiment.

4.2.2. *Results.* Participant responses in Experiment 2 are displayed in Tables 7–9. The number of times the first token is identified as the quotative is shown in Table 7 for questions that compare a token of the quotative with a token of the discourse particle and in Table 8 for questions that compare a token of the quotative with one of the traditionally grammatical functions of *like*. For questions comparing a traditionally grammatical function with the discourse particle, the number of times the first token is identified as the traditionally grammatical function is shown in Table 9.

For questions comparing the quotative with the discourse particle, roughly 54% of the questions where the first token is identified as the quotative have the quotative listed first on the answersheet. For questions comparing the quotative and one of the grammatical functions, approximately 52% of the 892 for which the first token is identified as the quotative are questions where the quotative context is listed first on the answersheet. Of the 1648 questions answered that compare the discourse particle with one of the traditionally gram-

Table 7. *Characteristics of quote–dp questions in Experiment 2 where the first token is identified as the quotative, by whether the participant is in a CR or a NCR group.*

feature questions comparing	CR girl quote–dp	NCR girl quote–dp	CR and NCR quote–dp
total number speakers	23	19	42
total questions answered	906	744	1650
total 1st token labeled as quote	456	375	831
quote first on sheet	250	195	445
1st token mean EucD	1.5330	1.5250	1.5290
2nd token mean EucD	1.5150	1.5820	1.5450
mean EucD diff. (Bark)	0.01807	−0.05693	−0.01577
1st token mean nuc F2 (Bark)	11.59	11.60	11.59
2nd token mean nuc F2 (Bark)	11.57	11.65	11.61
mean nuc F2 diff. (Bark)	0.01299	−0.04582	−0.01355
1st token mean duration ratio	0.3178	0.3198	0.3187
2nd token mean duration ratio	0.3406	0.3281	0.3350
mean duration ratio diff.	−0.022830	−0.008376	−0.016310
1st token [k] present, 2nd token [k] absent	58	49	107
1st token [k] absent, 2nd token [k] present	69	46	115
[k] present for both tokens	236	205	441
[k] absent for both tokens	93	75	168

Table 8. *Characteristics of quote–gram questions in Experiment 2 where the first token is identified as the quotative, by whether the participant is in a CR or a NCR group.*

feature questions comparing	CR girl quote–gram	NCR girl quote–gram	CR and NCR quote–gram
total number speakers	23	19	42
total questions answered	905	747	1652
total 1st token labeled as quote	494	398	892
quote first on sheet	234	183	462
1st token mean EucD	1.4480	1.4700	1.4580
2nd token mean EucD	1.4800	1.4550	1.4690
mean EucD diff. (Bark)	−0.03204	0.01479	−0.01114
1st token mean nuc F2 (Bark)	11.45	11.44	11.44
2nd token mean nuc F2 (Bark)	11.37	11.38	11.37
mean nuc F2 diff. (Bark)	0.08333	0.05636	0.07130
1st token mean duration ratio	0.3845	0.3853	0.3849
2nd token mean duration ratio	0.4194	0.4097	0.4151
mean duration ratio diff.	−0.03488	−0.02442	−0.03021
1st token [k] present, 2nd token [k] absent	131	102	233
1st token [k] absent, 2nd token [k] present	208	162	370
[k] present for both tokens	131	115	246
[k] absent for both tokens	24	19	43

Table 9. *Characteristics of gram–dp questions in Experiment 2 where the first token is identified as the grammatical function, by whether the participant is in a CR or a NCR group.*

feature questions comparing	CR girl gram–dp	NCR girl gram–dp	CR and NCR gram–dp
total number speakers	23	19	42
total questions answered	910	738	1648
total 1st token labeled as quote	441	366	807
quote first on sheet	248	202	450
1st token mean EucD	1.9500	1.9290	1.9400
2nd token mean EucD	1.9180	1.9580	1.9360
mean EucD diff. (Bark)	0.0316	−0.02898	0.004128
1st token mean nuc F2 (Bark)	11.45	11.47	11.46
2nd token mean nuc F2 (Bark)	11.54	11.48	11.51
mean nuc F2 diff. (Bark)	−0.08344	−0.01363	−0.05178
1st token mean duration ratio	0.4406	0.4398	0.4402
2nd token mean duration ratio	0.4321	0.4389	0.4352
mean duration ratio diff.	0.008490	0.0008953	0.005046
1st token [k] present, 2nd token [k] absent	88	66	154
1st token [k] absent, 2nd token [k] present	87	76	163
[k] present for both tokens	238	197	435
[k] absent for both tokens	28	27	55

Table 10. *Experiment 2 coefficients of fixed effects for model comparing responses to the quotative and the discourse particle.*

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.16468	0.09152	1.799	0.0720
quote second	−0.29861	0.12929	−2.310	0.0209
duration ratio difference	−0.55828	0.27063	−2.063	0.0391

matical functions, participants identify the first token as the traditionally grammatical function 807 times. Roughly 56% of these have the traditionally grammatical function listed first on the answersheet.

Three mixed effects models are fit to the data in order to investigate each of these factors within the context of the other phonetic factors investigated. Participant and question number are included as random effects in each model. Tested as potential fixed effects were a number of factors comparing phonetic cues from the first token with those in the second. These include the difference in /l/ to vowel duration, the difference in formant values at the nucleus target, the difference in Euclidean distance, and the difference in vowel duration. Also tested was the context listed first on the answersheet. Only factors that reach significance are included as fixed effects in the models and only these factors are listed in the tables shown.

The first model for Experiment 2 compares responses to the quotative and the discourse particle. The coefficients for the fixed effects are shown in Table 10.

Whether the quotative context appears first or second on the page significantly predicts responses; participants are less likely to identify the first token as the quotative if the quotative context is listed second ($p < 0.05$). This indicates a bias toward identifying the first auditory token played with the first context listed on the answersheet. As with results from Experiment 1, the other factor included in the model is significant above and beyond the effect from the context-order bias.

The other factor included in the model is the difference in the /l/ to vowel duration ratio between the first and second tokens (duration ratio difference). If the difference is positive, indicating that the /l/ to vowel duration ratio in the first token is larger than that in the second token, the first token is less likely to be identified as the quotative ($p < 0.05$). This result is consistent with trends from production. In production, tokens with a longer /l/ to vowel duration ratio are less likely to be the quotative than the discourse particle and, in perception, listeners are less likely to identify tokens with a longer /l/ to vowel duration ratio as the quotative. This result is also consistent with results from the first model from Experiment 1 which also compares responses to questions with the quotative and discourse particle as stimuli. Observing this result in

Table 11. *Experiment 2 coefficients of fixed effects for model comparing responses to the discourse particle and grammatical functions of* like.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−0.0416868	0.0756207	−0.5513	0.5815
F2 difference	−0.0006476	0.0003086	−2.0983	0.0359

Experiment 2 provides evidence that the trend observed in Experiment 1 is not due to phonetic cues in the preceding context but durational cues within the token of *like* itself.

A second mixed effects model was fit to the data from Experiment 2, comparing responses to questions where the stimuli are the discourse particle and one of the traditionally grammatical functions of *like*. The same factors that were tested in the first model were tested in the second model and only those that reach significance are included in the model shown in Table 11.

The only fixed effect included in the model is the difference between the F2 value at the nucleus target of the first and second token (F2 difference). When the value is positive (indicating that the first token has a greater F2 or a fronter nucleus), participants are less likely to identify the first token as the traditionally grammatical function ($p < 0.05$). In other words, tokens with a greater F2 are more likely to be matched with the discourse particle context and tokens with a smaller F2 are more likely to be matched with the context for the traditionally grammatical function. This tendency is consistent with production, where tokens of the traditionally grammatical function are more likely to be realized with a smaller F2 value at the nucleus target than are tokens of the discourse particle. This provides evidence that listeners are able to access stored information regarding the distribution of phonetic features across different lemmas that share a wordform.

The third mixed model was fit to the responses to questions comparing a quotative and a traditionally grammatical function of *like*. It models the likelihood that a token is the quotative. All of the factors that were tested in the first two models were also tested in the third model.

As shown in Table 12, the only factor reaching significance (and therefore the only fixed effect included in the model) is the difference between the F2 value of the first token's nucleus target and the F2 value of the second token's

Table 12. *Experiment 2 coefficients of fixed effects from the model comparing responses to the quotative and grammatical functions of* like.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.1724920	0.0954958	1.806	0.0709
F2 difference	0.0011511	0.0005074	2.269	0.0233

nucleus target (F2 difference). A positive difference in F2 indicates that the first token has a higher F2 at the nucleus target than the second token. As indicated by the positive coefficient for the factor in Table 12, the greater the difference in F2, the more likely a participant is to match the first token with the quotative context and the second token with the context for the traditionally grammatical function ($p < 0.05$). Again, this is consistent with trends in production because speakers are more likely to produce tokens of the quotative with a fronter nucleus target than tokens of traditionally grammatical functions.

4.3. *Experiment 3*

4.3.1. *Method.* For the third experiment, the questions of interest are whether participants can use fine phonetic detail to identify social information, namely, the lunch habits of the speaker and whether this interacts with token function. The experiment is divided into three blocks. In the first block, participants are told that they will hear the word *like*, as in the sentence *I like toast.* In the second block, they are told they will hear *like*, as in the sentence *I was like, "Yeah okay"*. In the third block, they are told that they will hear two tokens of *like* produced by the same speaker. They are told that the first token will have the same meaning as in the first block and the second token will have the same meaning as in the second block. The different blocks are used to expose participants to the different functions types at different times. Blocking by function simplifies the task, increasing the chance that participants will use additional information (e.g. their (not necessarily conscious) knowledge of sociophonetic trends) to identify the speaker. The participants are not misled; the actual auditory tokens themselves are either the quotative or the traditionally grammatical function, depending on the block.

Upon hearing a token of *like* in each of these blocks, participants are asked to indicate whether they believe the speaker eats lunch in the CR or not. They are then asked whether they recognize the speaker and, if so, who they believe the speaker to be. The stimuli include voices from 10 girls (5 CR, 5 NCR). The same voices are used in each of the blocks, though the voices are presented in different orders within each block. The tokens used in Block 3 are the same as those used in Blocks 1 and 2.

4.3.2. *Results.* The number of times a voice in Block 1 is (correctly or incorrectly) identified as someone who eats lunch in the CR is shown in Table 13 for CR and NCR participants. Of all 1224 responses, participants indicate that they recognize the speaker 236 times, or 19.3% of the time, and fail to answer the question of whether they recognize the speaker 17 times. The low

Table 13. *Characteristics of the grammatical functions in Experiment 3, Block 1 for questions where the voices are identified as someone who eats lunch in the CR. The total possible based only on questions answered is shown in parentheses.*

feature	CR subjects	NCR subjects	CR and NCR subjects
total number speakers	23	19	42
total questions identified as CR	151 (228)	111 (182)	262 (410)
actual voice = CR	87 (114)	56 (90)	143 (204)
recognize voice	41 (46)	25 (37)	66 (83)
mean EucD (Bark)	1.63000 (1.62400)	1.69600 (1.62200)	1.65800 (1.62300)
mean duration ratio	0.3772 (0.3770)	0.3725 (0.3760)	0.3752 (0.3766)
[k] realized	125 (183)	88 (145)	213 (328)
[lai]	26 (45)	23 (37)	49 (82)
[laik]	125 (183)	88 (145)	213 (328)

percentage of times that the speaker is recognized (correctly or incorrectly) suggests that participants identify a speaker as someone who eats lunch in the CR or not without necessarily having a particular speaker in mind. Of the 236 times they indicate recognizing the speaker, participants identify the girl who they believe produced the utterance 198 times. Of these, they correctly identify the speaker 53 times (26.8%) and someone who is not the speaker but the speaker's good friend 14 times (7.1%).

All stimuli in Block 1 are tokens of a traditionally grammatical function of *like*. For roughly 64% of the 410 questions answered, participants respond that the voice belongs to a girl who eats lunch in the CR. Questions in Block 2 only include tokens of quotative *like* as stimuli. As in the first block, there is an overall bias toward indicating that the speaker eats lunch in the CR. The raw counts of responses that the speaker is a girl who eats lunch in the CR is shown in Table 14.

For each question in Block 3, tokens of both a traditionally grammatical function and the quotative are played. As in the first two blocks, there is an overall bias toward responding that the speaker is someone who eats lunch in the CR; voices are identified as someone who eats lunch in the CR roughly 64% of the time. As with the first two experiments, apparent tendencies evident in the raw data should be viewed with some caution. There is a danger that some apparent trends can be attributed to correlations of the different phonetic characteristics in the stimuli or an artifact of response biases, such as the tendency to indicate that the voice is a girl who eats lunch in the CR.

In all three blocks, a voice is identified as a CR girl more often if the participant indicates recognizing it. In a mixed effects model fit to all the data from Experiment 3 and predicting responses on the task, this tendency for participants who recognize the voice to identify the speaker as a CR girl is highly significant, irrespective of whether they identify the correct girl or if the voice

Table 14. *Characteristics of the quotatives in Experiment 3, Block 2 for questions where the voices are identified as someone who eats lunch in the CR. The total possible based only on questions answered is shown in parentheses.*

feature	CR subjects	NCR subjects	CR and NCR subjects
total questions identified as CR	159 (211)	116 (188)	275 (399)
actual voice = CR	78 (106)	58 (94)	136 (200)
recognize voice	18 (24)	30 (40)	48 (64)
mean EucD (Bark)	1.2800 (1.2720)	1.2250 (1.2620)	1.2570 (1.2680)
mean duration ratio	0.3478 (0.3490)	0.3199 (0.3459)	0.3361 (0.3475)
[k] realized	70 (84)	52 (75)	122 (159)
[k] dropped	89 (127)	64 (113)	(240)
[la]	16 (21)	16 (19)	32 (40)
[lak]	18 (21)	14 (19)	32 (40)
[lai]	74 (106)	48 (94)	122 (200)
[laik]	52 (63)	38 (56)	90 (199)

is actually someone who eats lunch in the CR ($p < 0.0001$). Participants identify the speaker correctly approximately 59% of the time. While this is above chance, it suggests that familiar-sounding voices are identified as CR girls. Because voice-familiarity appears to have influenced responses so strongly, the remainder of the analysis is conducted only on data where participants indicate that they do not recognize the voice.

Responses to the grammatical functions in Block 1 were tested, but no phonetic factors appear to affect responses. The analysis presented here focuses on responses to questions in Blocks 2 and 3, all of which contain a token of quotative *like*. A binomial mixed effects model with question number and participant as random effects is fit to responses from Experiment 3 for only those questions that contained a token of quotative *like*. It models the likelihood of indicating on the answersheet that the speaker eats lunch in the CR. This is done in order to determine whether participants rely on particular phonetic cues in the voice when deciding whether the speaker is a CR girl and, if so, whether these phonetic cues reflect the phonetic variation observed in the production data.

In fitting the model, a number of factors were tested. These include whether the stimulus has a vowel that is monophthongal or a /k/ that is realized and whether the participant and stimulus voice are CR girls. Only factors that reach significance are included as fixed effects in the model. These are a combination of whether the quotative token in the stimulus has the /k/ present and is monophthongal. Thus, there are four factors in this factor group: one for tokens with a monophthongal vowel and /k/ dropped ([la]), one for tokens with a monophthongal vowel and /k/ present ([lak]), one for tokens with a diphthongal vowel and /k/ dropped ([lai]), and one for tokens with a diphthongal vowel and /k/ realized ([laik]). The coefficients for the model are shown in

Table 15. *Differences between the grammatical function and the quotative in Experiment 3, Block 3 for questions where the voices are identified as someone who eats lunch in the CR. The total possible based only on questions answered is shown in parentheses.*

feature	CR subjects	NCR subjects	CR and NCR subjects
total questions identified as CR	135 (214)	118 (184)	253 (398)
actual voice = CR	72 (108)	57 (92)	129 (200)
recognize voice	35 (42)	35 (47)	70 (89)
voice = extroverted	83 (128)	73 (111)	156 (239)
mean EucD diff. (Bark)	1.3440 (1.2280)	1.2240 (1.2930)	1.2880 (1.2580)
mean duration ratio diff.	0.31020 (0.32450)	0.28800 (0.31860)	0.29990 (0.32170)
gram. [k] realized, quote [k] not realized	78 (130)	63 (110)	141 (240)
gram. [k] not realized, quote [k] realized	27 (42)	26 (37)	53 (79)
both [k] realized	30 (42)	29 (37)	59 (79)
gram. [laik] quote [la]	17 (22)	14 (17)	31 (39)
gram. [laik] quote [lak]	15 (21)	13 (19)	28 (40)
gram. [laik] quote [lai]	61 (108)	49 (93)	110 (201)
gram. [laik] quote [laik]	15 (21)	16 (18)	31 (39)
gram. [lai] quote [laik]	27 (42)	26 (37)	53 (79)

Table 16. The tokens with a diphthongal vowel and the /k/ dropped are the default in the model and have estimated coefficient of zero.

Quotative tokens that are diphthongal and have the /k/ dropped are significantly less likely to be identified as having been produced by a CR girl than any of the other tokens ([la], $p < 0.01$; [laik], $p < 0.001$; [lak], $p < 0.05$). There is no significant difference in response between any of the other three realizations in the stimuli. The token with both a monophthongal vowel and the /k/ dropped ([la]) is most likely to be identified as having been produced by a CR girl.

Interestingly, the trends evident in Table 16 are carried almost entirely by the NCR girls; when the same model is run only on the subset of data from CR girls, none of the factors reach significance and the factors are listed in a

Table 16. *Coefficients of fixed effects for Experiment 3. Higher coefficients indicate a higher probability of identifying the speaker as someone who eats lunch in the CR.*

	Estimate	Std. Error	z value	Pr(> |z|)
(Intercept)	0.2174	0.2070	1.050	0.293692
quote = la	1.0300	0.3909	2.635	0.008422
quote = laik	0.8787	0.2526	3.479	0.000504
quote = lak	0.7922	0.3853	2.056	0.039784

different order in terms of their coefficients ([laik] > [lak] > [la] > [lai]). In contrast, when the same model is run only on the subset of data from NCR girls, all of the factors maintain the same level of significance and are in the same order in terms of their coefficients ([la] > [laik] > [lak] > [lai]) as when the model is run on all of the data. This interaction between eating place and phonetic realisation does not reach significance if included in the model shown in Table 16. This may be because, for both groups of speakers, tokens realised as [lai] are least likely to be identified as having been produced by a CR girl, and the trend for the CR participants was not robust. These findings are discussed further in the following section.

5. Discussion

Results from the first two experiments provide evidence that listeners can use phonetic information inherent in the stimuli to identify a token's function, even when distinguishing between different lemmas that share a wordform. That they do so in a manner that is consistent with some of the trends observed in production suggests that rich phonetic information, such as the /l/ to vowel duration ratio, is stored in the mind, indexed to lemma-based information, and accessed during speech perception. The results from Experiment 1 also provide evidence that individuals store information about the surrounding context. Quotative *like* is most frequently found in the first person and in the habitual present, and participants are more likely to identify a token as quotative *like* if the preceding context is in the first person or habitual present than if it is in, for example, the third person or the past tense.

These findings are consistent with an exemplar-based model of speech perception and production in which utterances are stored in the mind complete with fine-grained phonetic detail and indexed with contextual information observed at the time of the utterance (Pierrehumbert 2001, 2002). The results presented here indicate that such information must include the grammatical function of a token.

In Experiment 3, participants are more likely to indicate that the speaker is a CR girl if they believe they recognize the voice, even if they incorrectly identify the speaker ($p < 0.0001$). In other words, if a voice sounds familiar, the speaker is (correctly or incorrectly) identified as someone who eats lunch in the CR. This is not entirely surprising, as CR girls are involved in more school activities and are more visible at the school. They are talkative in class, they play sport, and they have leadership roles. With few exceptions, NCR groups interact with each other rarely, and some are actually more likely to interact with CR girls than with someone from a NCR group other than their own. Therefore, a wider variety of students have exposure to CR girls' speech. CR girls have less exposure to NCR girls (and their speech) than to other CR girls, and NCR girls have more exposure to CR girls (and their speech) than to girls from other NCR groups.

This may also explain why NCR girls appear to be sensitive to some of the phonetic cues in the stimuli but CR girls are not. NCR participants were least likely to identify tokens with a diphthong and the /k/ dropped ([lai]) as having been produced by a CR girl when compared with all other realizations present in the stimuli. In terms of diphthongization, this is consistent with results from production: CR girls are more likely to produce a more monophthongal vowel in realizations of all of the different functions of *like* than NCR girls. Comparing the two stimulus realizations where the /k/ is dropped ([lai] and [la]), NCR girls appear to be sensitive to the trend from production when identifying the eating place of the speaker. That no trends arise in the data from CR girls may be because they have less exposure to the speech of NCR girls and therefore may not be as sensitive to differences between the two groups. But why, among NCR participants, is there no significant difference between stimuli with monophthongal vowels and those with diphthongal vowels when the /k/ is present? And why don't the listeners identify more of the stimuli with /k/ present as having been produced by girls who do not eat lunch in the CR?

The standard realization of *like* has a diphthong and a realized /k/ ([laik]), as evidenced by the shift made by the girls while reading the production task; all girls, even those who frequently drop the /k/ in spontaneous speech, produce a diphthong and a /k/ in all tokens during the reading task. CR girls are viewed as "normal" at the school and it is possible that this could bias individuals toward identifying a greater number of tokens with the standard realization of *like* as having been produced by "normal" CR girls, despite the fact that NCR girls are more likely to produce a diphthong and to realize the /k/ in quotative *like*. In terms of the other three realizations, if CR girls drop the /k/ and produce a monophthong in quotative *like* and NCR girls realize the /k/ and produce a diphthong, then stimuli realized as [lai] and [lak] have conflicting information as to which group the speaker might belong: is [lak] produced by a CR girl because the vowel is monophthongal, or by a NCR girl because the /k/ is realized? This conflict in speaker information may have affected responses. Interestingly, NCR participants are most likely to identify the voice as someone who eats lunch in the CR if the stimulus is unambiguously consistent with CR girls' production (the tokens have a monophthongal vowel and a dropped /k/ ([la])), though the only difference that reaches significance is with [lai]. While the results provide some evidence that perceivers can extract social information based on phonetic cues, whether this information can be extracted in lemma-specific ways remains inconclusive.

Taken together, these results provide evidence that perceivers are sensitive to the relationship between phonetic and lemma-based information during perception. In production, phonetic variation depends on the social group of the individual and the function of the token. In perception, individuals are sensitive to the relationship between phonetic and lemma-based information. They also extract community-specific social information about the speaker, depending on whether the voice sounds familiar and whether the token of quotative *like* is realized as [lai]. This suggests that social, phonetic, and grammatical (syntactically/semantically-defined) information is stored in, or indexed to, the lexicon and can be accessed during the perception of speech. However, if perceivers' notions of "normal" and "different" affect the extraction of social information from auditory input, then language ideology must also play some role in perception, even if it does not accurately reflect patterns that are actually present in the data.

6. Conclusion

In sum, the results from these experiments provide evidence that individuals are sensitive to grammatical (lemma-based) variation during perception; very subtle grammatical variability must be indexed to fine-grained phonetic detail in the mind. The results also provide evidence that the storage of sociophonetic relationships is complex; speech perception models need to be developed so that they include indexation between stored phonetic information and a listener's beliefs regarding language and the language ideology of a speech community.

Correspondence e-mail address: katiedrager@gmail.com

Notes

1. There is some evidence that for Māori English speakers, quotative *like* is more likely to be produced in the past tense than in the habitual present D'Arcy (2008). Because the vast majority of the participants are speakers of Pākehā English, I will use the terms "likely" and "unlikely" to refer to the organization of the stimuli, although these terms would not be appropriate for an ethnicity-based investigation.
2. Interestingly, the Māori English speakers who participated in the experiment responded in the opposite direction from the Pākehā participants with regard to this factor. This is consistent with trends in the production of quotative *like* in Māori and Pākehā Englishes. Further work is needed to determine the extent to which perceivers from different social groups use lemma-specific contextual information that is consistent with socially-conditioned trends from production.

Acknowledgements

I would like to thank my PhD supervisor, Jennifer Hay, for her help with this project. I am also grateful to Alexandra D'Arcy and Robert Fromont. This paper has benefited from comments made by Benjamin Munson, Keith Johnson, members of the socio-kaffeeklatsch at the University of Canterbury, the audience at the 11th Laboratory Phonology conference, and two anonymous reviewers. I would also like to thank Selwyn Girls' High; I am especially grateful to the students, who took part in interviews and experiments and who allowed me a window into their lives.

References

Anderson, Gisle. 2001. *Pragmatic Markers and Sociolinguistic Variation: A relevance-theoretic approach to the language of adolescents*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Baayen, R. Harald. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics*. Cambridge: Cambridge University Press.

Bentler, P.M., Douglas N. Jackson & Samuel Messick. 1971. Identification of content and style: A twodimensional interpretation of acquiescence. *Psychological Bulletin* 76(3). 186–204.

Buchstaller, Isabelle & Alexandra D'Arcy. 2009. Localized globalization: A multi-local, multivariate investigation of quotative 'be like'. *Journal of Sociolinguistics* 13(3). 291–331.

Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14. 261–290.

Campbell-Kibler, Kathryn. 2007. Accent, (ING), and the social logic of listener perceptions. *American Speech* 82. 32–64.

Dailey-O'Cain, Jennifer. 2000. The sociolinguistic distribution of and attitudes toward focuser *like* and quotative *like*. *Journal of Sociolingistics* 4(1). 60–80.

D'Arcy, Alexandra. 2007. *Like* and language ideology: Disentangling fact from fiction. *American Speech* 82(4). 386–419.

D'Arcy, Alexandra. 2008. Quoting ethnicity: Constructing dialogue in Aotearoa. In: *Methods in Dialectology 13*, University of Leeds.

Drager, Katie. 2009a. Language, stance, and identity at Selwyn Girls High. In: *Proceedings of the 5th International Gender and Language Association Conference*. Wellington: Victoria University of Wellington.

Drager, Katie. 2009b. *A Sociophonetic Ethnography of Selwyn Girls' High*. New Zealand: Unpublished PhD thesis from the University of Canterbury.

Eckert, Penelope. 1989. *Jocks and Burnouts: Social categories and identity in the high school*. Teachers College Press.

Fromont, Robert & Jennifer Hay. 2008. ONZE Miner: the development of a browser-based research tool. *Corpora, 3*(2), 173–193.

Gahl, Susanne. 2008. 'Time' and 'thyme' are not homophones: The effect of lemma frequency on word durations in a corpus of spontaneous speech. *Language* 84(3). 474–496.

Hay, Jen & Joan Bresnan. 2006. Spoken syntax: The phonetics of giving a hand in New Zealand English. *The Linguistic Review* 23(3). 321–349.

Hay, Jen & Katie Drager. 2007. Sociophonetics. *Annual Review of Anthropology* 36. 89–103.

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4). 458–484.

Hazen, Kirk. 2006. The final days of Appalachian heritage language. In: Thomas E. Murray & Beth L. Simon (eds.), *Language Variation and Change in the American Midland: a new look at Heartland English*, 129–150. Amsterdam: John Benjamins Publishing Company.

Jescheniak, Jorg D. & Willem J. M. Levelt. 1994. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology* 20(4). 824–843.

Johnson, Keith. 1997b. Speech perception without speaker normalization. In: K. Johnson & J. Mullenix (eds.), *Talker variability in speech processing*, 146–165. San Diego: Academic Press.

Jurafsky, Daniel, Alan Bell & Cynthia Girand. 2002. The role of lemma in form variation. In: Carlos Gussenhoven & Natasha Warner (eds.), *Papers in Laboratory Phonology VII*, 1–34. Berlin/New York: Mouton de Gruyter.

Levelt, W., A. Roelofs & A. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22. 1–75.

Maegaard, Marie. to appear. Linguistic practice and stereotypes among Copenhagen adolescents. In: Pia Quist & Bente Ailin Svendsen (eds.), *Multiethnolects and Linguistic Practices in Multiethnic Urban Scandinavia*. Bristol: Multilingual Matters.

Munson, Benjamin & Nancy Pearl Solomon. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 5(5). 1048–1058.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In: Joan Bybee & Paul J. Hopper (eds.), *Frequency Effects and Emergent Grammar*, 137–158. Amsterdam: John Benjamins.

Pierrehumbert, Janet. 2002. Word-specific phonetics. In: Carlos Gussenhoven & Natasha Warner (eds.), *Papers in Laboratory Phonology VII*, 101–139. Berlin/New York: Mouton de Gruyter.

Pierrehumbert, Janet. 2006. The next toolkit. *Journal of Phonetics* 34. 516–530.

Plug, Leendert. 2005. Phonetic reduction and categorisation in exemplar-based representation: Observations on a Dutch discourse marker. In: *Proceedings of ConSOLE XIII*.

R Development Core Team. 2007. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Romaine, Suzanne & Deborah Lange. 1991. The use of *like* as a marker of reported speech and thought: A case of grammaticalization in progress. *American Speech* 66(3). 227–279.

Strand, Elizabeth. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18(1). 86–99.

Strand, Elizabeth & Keith Johnson. 1996. Gradient and visual speaker normalization in the perception of fricatives. In: D. Gibbon (ed.), *In Natural Language Processing and Speech Technology*, 14–26. Berlin: Mouton.

Tagliamonte, Sali & Rachel Hudson. 1999. *Be like* et al. beyond America: The quotative system in British and Canadian youth. *Journal of Sociolinguistics* 3(2). 147–172.

Tagliamonte, Sali A. & Alexandra D'Arcy. 2007. Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change* 19. 199–217.

Zipf, George K. 1929. Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 15. 1–95.